Sleeping Beauty in a Grain of Rice

Sleeping Beauty in a Grain of Rice

David Haig

Department of Organismic and Evolutionary Biology,

Harvard University, 26 Oxford Street,

Cambridge MA 02138.

phone: 1-617-496-5125

fax: 1-617-495-5667

e-mail: dhaig@oeb.harvard.edu

Acknowledgments Lucas Mix brought the Sleeping Beauty problem to my attention. Carl Veller patiently explained thirder reasoning and critically read the manuscript. Ned Hall and the anonymous reviewers provided valuable input.

Abstract In the Sleeping Beauty problem, Beauty is woken once if a coin lands heads or twice if the coin lands tails but promptly forgets each waking on returning to sleep. Philosophers have divided over whether her waking credence in heads should be a half or a third. Beauty has centered beliefs about her world and about her location in that world. When given new information about her location she should update her worldly beliefs before updating her locative beliefs. When she conditionalizes in this way, her credence in heads is a half before and after being told it is Monday. In applications of Dutch Book arguments to the Sleeping Beauty problem, the probability of a particular outcome has often been confounded with consequences of that outcome. Heads and tails are equally likely but twice as much is at stake if the coin falls tails because Beauty is fated to make the same choice twice. As a consequence, the possibility of tails should be given twice the weight of the possibility of heads when deciding whether to bet on heads even though heads and tails are equally likely.

Keywords Sleeping Beauty; Hamilton's rule; credence; relatedness; endosperm; conditionalization; *de se* beliefs

Elga (2000) introduced the Sleeping Beauty problem as a paradigm for thinking about centered beliefs. In this puzzle, Beauty is uncertain whether a fair coin landed heads or tails. If the coin landed heads then she will be woken once. If the coin landed tails, then she will be woken twice. Beauty is assumed to understand the procedure but to have all memories of waking erased on returning to sleep. What should be her waking credence that the coin landed heads? 'Thirders' believe the answer is one-third because Beauty is woken twice as often after the coin lands tails as after heads (Elga 2000; Hitchcock 2004;

Briggs 2010). 'Halfers' believe the answer is one-half because the coin is equally likely to land heads or tails (Lewis 2001; Arntzenius 2002; Meacham 2008). Halfers and thirders continue to lock swords at the time of writing.

The purpose of this paper is to offer a resolution of the Sleeping Beauty problem informed by consideration of a parallel problem in botany that arose in the 1980s in which some theoreticians adopted a 'thirder' stance and others a 'halfer' stance. This problem concerned the genetic constitution of endosperm, a tissue within seeds. My understanding of the philosophers' problem was clarified by thinking about the botanical problem and vice versa. Philosophical disagreement between halfers and thirders may seem abstruse to empirical scientists and the genetics of endosperm may seem equally omphaloscopic to philosophers until it is noted that every grain of rice, wheat, rye, oats, barley, millet, sorghum, and maize is predominantly endosperm. Most calories in the global human diet come from the direct consumption of endosperm or its indirect consumption via eating grain-fed beasts. Epistemology can be practical.

I propose that Sleeping Beauty should update her beliefs about her world before updating her beliefs about her location in that world and when she does this she maintains a consistent credence of one half in heads. The next two sections, *Eternal Beauty* and *Ephemeral Beauty*, present the argument which is summarized in *Beauty at Rest*. *Gambling on Beauty* offers an explanation of why arguments based on the bets Beauty would accept on particular outcomes have often appeared to support the thirder position. *Hamilton's Wager* and subsequent sections present the Sleeping Beauty problem as instantiated within a grain of rice.

Eternal Beauty

Consider an infinite variant of the Sleeping Beauty problem. Eternal Beauty is told that she will be put to sleep and woken on every Monday for eternity if a fair coin lands heads but on every Monday and Tuesday for eternity if the coin lands tails. On waking she will be told neither the day nor the outcome of the coin toss, and she will forget each and every waking on returning to sleep. Before being put to sleep for the first time, she understands the protocol and she retains her understanding of the protocol at each awakening. She believes in three possibilities on waking: the coin landed heads and this is Monday (H_1), the coin landed tails and this is Monday (T_1), or the coin landed tails and this is Tuesday (T_2).

Eternal Beauty's beliefs about her world should be distinguished from her beliefs about her location in her world. I will call the former her worldly beliefs (P') and the latter her locative beliefs (P).[1] She believes P'(heads) = P'(tails) = ½. If the coin lands heads, then her world contains H_1, P'(H_1) = ½. If the coin lands tails, then her world contains T_1 and T_2, P'(T_1) = P'(T_2) = ½. P'(T_1) and P'(T_2) are duplicates of the probability of her possible world in which the coin landed tails. Eternal Beauty knows she will wake on Mondays, P'(H_1) + P'(T_1) = 1, and believes she has a half chance of waking on Tuesdays, P'(T_2) = ½.

[1] Locative beliefs are *de se* beliefs. Worldly beliefs are not otherworldly beliefs. They are Beauty's centered beliefs about her actual world and what it might be. If her actual world is conceived as an object with properties, then wordly beliefs might be considered *de re* (but I am ill-educated on the philosophical nuances of these Latin phrases).

How should Eternal Beauty convert beliefs about her world into beliefs about her location in her world? If she is in a world with T_1 wakings then she must also be in a world with T_2 wakings, namely $P'(T_1 \mid T_2) = P'(T_2 \mid T_1) = 1$. However, T_1 and T_2 are mutually exclusive when she wakes, $P(T_1 \mid T_2) = P(T_2 \mid T_1) = 0$. She believes her world contains either H_1 or (T_1 and T_2) but, on waking, is located in H_1 or T_1 or T_2. How should she conceptualize this peculiar transformation of (T_1 and T_2) into (T_1 or T_2)?

Halfers believe that waking on Monday is twice as likely in possible worlds created by the coin landing tails as in possible worlds created by the coin landing heads. Eternal Beauty reasons on waking that either the coin landed heads or the coin landed tails. If the coin landed heads, with probability P(heads) = ½, then it is Monday (H_1), but if the coin landed tails, with probability P(tails) = ½, then it is either Monday (T_1) or Tuesday (T_2). Since $P(T_1) = P(T_2)$, by a principle of indifference, Eternal Beauty's locative beliefs on waking are $P(H_1) = ½$, $P(T_1) = ¼$, $P(T_2) = ¼$.

Thirders believe that Eternal Beauty is woken twice as often in possible worlds created by the coin landing tails as in possible worlds created by the coin landing heads. As a corollary thirders believe that Eternal Beauty is woken on Monday just as often when the coin lands heads as when the coin lands tails. Because H_1, T_1, and T_2 occur equally often, she believes $P(H_1) = P(T_1) = P(T_2) = 1/3$ and, therefore, P(heads) = 1/3.

For halfers, T_1 and T_2 collectively have the same locative probability as H_1 on waking but, for thirders, T_1 and T_2 individually have the same locative probability as H_1. Halfers believe, and thirders probably agree, that the likelihood that a waking is on Monday in possible worlds in which the coin lands heads is twice the corresponding likelihood in possible worlds in which the coin lands tails. Thirders also believe that waking on Monday occurs *as often* in her possible world in which the coin lands heads

as in her possible world in which the coin lands tails. In reasoning across these possible worlds, halfers use the *likelihood* of Monday but thirders use the *frequency* of Monday.

What if you asked Eternal Beauty about her beliefs? Perhaps she would say that whether her world was created heads or tails, whether it is Monday or Tuesday, or whether there are any other days but this day, are metaphysical questions because there is nothing she can learn to distinguish among the alternatives. She has a memory that P'(heads) = ½ from before the procedure—was it yesterday?—but lives in an eternal present. Why should she have prior beliefs if nothing is at stake? Possible worlds in which the coin toss was heads or the coin toss was tails are both infinite sets of indistinguishable days, without past or future, in eternal recurrence. If she were a number theorist, Beauty might consider the proposition 'one countable infinite set has twice as many members as another' to be meaningless.

Ephemeral Beauty

Eternal Beauty's sister, Ephemeral Beauty, is told on Sunday that a fair coin will be tossed. If it comes up heads she will be woken on Monday, and then made to forget, but if it comes up tails she will be woken on Monday and Tuesday, and made to forget after each waking. She will then be woken and debriefed on Wednesday. This is the original Sleeping Beauty problem. The halfer can simply argue that, when Ephemeral Beauty is woken on Monday or Tuesday, her beliefs are the same as she possessed before the procedure. She believes P(heads) = ½ on Sunday before the coin was tossed and believes the same on Wednesday before the outcome of the coin toss is revealed. It would be perverse for her to believe anything different on Monday or Tuesday. A thirder must

argue that Ephemeral Beauty believes tails to be twice as likely as heads on Monday and Tuesday despite contrary beliefs on Sunday and Wednesday.

Elga (2000) and Lewis (2001) updated Ephemeral Beauty's beliefs using standard conditionalization but different priors. They agreed that P(heads) = $P(H_1)$, that $P(T_1) = P(T_2)$, and that when she is told it is Monday her credence in heads should increase because the prior possibility of T_2 is eliminated. This led each to adopt positions they found counterintuitive. Elga (2000) reasoned that P_{Mon}(heads) = ½ and working backward was forced to conclude that P(heads) = 1/3. By contrast, Lewis (2001) assumed P(heads) = ½ and by working forward was forced to conclude that P_{Mon}(heads) = 2/3. Both reasoned that, before being told it is Monday, Sleeping Beauty believes $P(H_1) + P(T_1) + P(T_2) = 1$ but, after being told it is Monday, she believes $P_{Mon}(H_1) + P_{Mon}(T_1) = 1$. The probability formerly attached to T_2 was distributed between H_1 and T_1. For Elga, Beauty's centered beliefs changed from $P(H_1) = P(T_1) = P(T_2) = 1/3$ to $P_{Mon}(H_1) = P_{Mon}(T_1) = ½$, $P_{Mon}(T_2) = 0$. For Lewis, Beauty's centered beliefs changed from $P(H_1) = ½$, $P(T_1) = ¼$, $P(T_2) = ¼$ to $P_{Mon}(H_1) = 2/3$, $P_{Mon}(T_1) = 1/3$, $P_{Mon}(T_2) = 0$.

Elga and Lewis conditionalized Ephemeral Beauty's locative beliefs after learning it is Monday from her prior locative beliefs before learning it is Monday. But what if Ephemeral Beauty first updated her worldly beliefs before updating her locative beliefs? On waking, her worldly beliefs are $P'(H_1) = P'(T_1) = P'(T_2) = ½$ but, on being told it is Monday, she learns her possible worlds for this day do not include Tuesday, $P'_{Mon}(T_2) = 0$. Therefore, $P'_{Mon}(H_1) = P'_{Mon}(T_1) = ½$ and $P_{Mon}(H_1) = P_{Mon}(T_1) = ½$. The duplicate probability of her worldly beliefs 'evaporates' when she is told it is Monday. The locative probability formerly attached to T_2 is transferred to T_1. Ephemeral Beauty is a 'double-halfer' who believes both P(heads) = ½ and P_{Mon}(heads) = ½ (Bostrom 2007;

Meacham 2008). The information it is Monday tells her nothing about the coin toss and does not change her credence in heads. [2]

The above procedure differs from standard conditionalization in how it handles duplicate probability as instantiated in $P'(T_1)$ and $P'(T_2)$. If new information eliminates one (but not all) of the duplicates, then her other worldly beliefs are unaffected: $P'(T_2) = ½$ conditionalizes to $P'_{Mon}(T_2) = 0$ but $P'(H_1)$ and $P'(T_1)$ remain unchanged. P'(tails) was formerly represented redundantly by $P'(T_1)$ and $P'(T_2)$ but is now represented solely by $P'_{Mon}(T_2)$. When worldly beliefs are converted to locative beliefs, P'(tails) = ½ is distributed among the uneliminated duplicates, in this case to the single remaining option $P(T_1) = ½$.[3]

All Beauty learns on being told that it is Monday is that it is not Tuesday. She learns nothing about the toss of the coin. A simple reframing of her prior beliefs may make this claim more intuitive: (i) Beauty believes she will be woken on Monday, P'(Monday) = 1; (ii) Beauty believes she will be woken on Tuesday only if the coin lands tails, P'(Tuesday | heads) = 0; and (iii) Beauty believes the coin is fair, P'(heads) = ½. When told it is Monday she learns that this is the first time she has woken but she learns

---

[2] Lewis (1979) might have said that when Beauty is told it is Monday, she learns something about her location in ordinary space that changes her location in logical space. Her *propositional attitude* changes from 'week in which heads or tails' to 'Monday in which heads or tails.' 'Waking on Tuesday' is a *property* of the first propositional attitude that does not have a counterpart in the second propositional attitude.

[3] This procedure appears similar to, perhaps is the same as, Meacham's (2008, p. 249) compartmentalized conditionalization.

nothing about whether she will wake the next day. That possibility is still in the future depending on an unknown flip of the coin.

Beauty at Rest

Beauty *knows* what she believes to be true. At each particular sentient moment, Beauty has beliefs about her *world* associated with worldly probabilities (P') and beliefs about her *location* in that world associated with locative probabilities (P). Worldly and locative probabilities may be primary probabilities, based on things she knows, or derivative probabilities, based on primary probabilities. On waking, Beauty knows the coin to be fair and believes that either the coin landed heads or the coin landed tails. P'(heads) = P'(tails) are her primary worldly probabilities of her possible worlds.

Beauty also believes that if the coin landed heads then she will wake on Monday (H_1) but if the coin landed tails she will wake on Monday (T_1) and Tuesday (T_2). $P'(H_1) = P'(T_1) = P'(T_2) = ½$, are her derived worldly probabilities where $P'(H_1)$ is derivative of P'(heads) and $P'(T_1)$ and $P'(T_2)$ are derivative of P'(tails). $P'(T_1)$ and $P'(T_2)$ are duplicate worldly probabilities. If the coin lands tails, both occur in Beauty's world. She knows that she wakes on Monday because $P'(H_1) + P'(T_1) = 1$, but believes she has a half chance of waking on Tuesday, because $P'(T_2) = ½$.

Beauty's locative beliefs have the same probability as her corresponding worldly beliefs except for duplicate derived probabilities in which case the primary probability is divided among the duplicates in locative beliefs. Thus for the non-duplicate probabilities $P'(\text{heads}) \rightarrow P(\text{heads})$, $P'(\text{tails}) \rightarrow P(\text{tails})$, $P'(H_1) \rightarrow P(H_1)$, but for the duplicate probabilities $P'(\text{tails}) \rightarrow P(T_1) = P(T_2) = ¼$. Beauty's worldly beliefs on waking are $P'(H_1) = P'(T_1) = P'(T_2) = ½$ and her locative beliefs are $P(H_1) = ½$, $P(T_1) = P(T_2) = ¼$. When

information is provided relevant to her locative beliefs, Beauty first updates P' before updating P. Thus, on being told it is Monday, Beauty updates her worldly beliefs, eliminating the possibility of Tuesday, and then uses her new worldly beliefs to update her locative beliefs:

$$\{P'_{Mon}(H_1) = ½, P'_{Mon}(T_1) = ½, P'_{Mon}(T_2) = 0\} \rightarrow \{P_{Mon}(H_1) = ½, P_{Mon}(T_1) = ½, P_{Mon}(T_2) = 0\}$$

$$P'_{Mon}(\text{heads}) = ½ \rightarrow P_{Mon}(\text{heads}) = ½.$$

Gambling with Beauty

An experimental economist remained unconvinced by such philosophical arguments. From his perspective, Ephemeral Beauty's beliefs are no less metaphysical than Eternal Beauty's beliefs if they have no material consequences. He commanded the research budget of an economist rather than a philosopher and proposed that the only way to understand credences is for subjects to have something at stake. For this purpose, he recruited many beauties to undergo the Sleeping Beauty procedure or a minor variant thereof. First, he assessed his recruits' beliefs prior to the procedure. He reasoned that, if the beauties were risk-neutral gamblers, they would accept bets on heads with payout B for stake C whenever $rB - C > 0$, or $C/B < r$, where r was their credence in heads. The economist found that the beauties accepted bets on heads for $C/B < ½$ but rejected bets for $C/B > ½$. He therefore concluded they believed P(heads) = ½.

All beauties were told that, each time they woke, they would be offered a series of bets on heads with different values of B and C to probe their beliefs about P(heads). Their stakes would be collected and their winnings paid on Wednesday. If a coin landed heads, then they would be woken on Monday. If the coin landed tails, then they would be woken on Monday and Tuesday. At each waking, they would be told neither the

outcome of the coin toss nor the day and their memories of waking would be erased on returning to sleep.

The beauties were then assigned to one of two groups. OR-beauties were told that, if the coin landed heads, their bets on Monday would be honored, but, if the coin landed tails, only their bets on Monday *or* Tuesday would be honored (with the choice of Monday or Tuesday determined by an independent toss of the same coin). AND-beauties were told that all bets would be honored. If the coin landed heads, their bets on Monday would be honored, but, if the coin landed tails, their bets on Monday *and* Tuesday would be honored.

OR-beauties used the same decision rule during the procedure as they used before the procedure, $C/B < ½$. They continued to believe P(heads) = ½ on waking. The entire experimental rigmarole could have been avoided. It changed nothing of relevance.

AND-beauties, by contrast, adopted the decision rule $C/B < 1/3$ during the procedure. They seemed to believe that P(heads) = 1/3. However, when the economist came to settle his accounts on Wednesday he realized he had misunderstood their pecuniary incentives. Heads and tails were equally likely. For each accepted wager, he paid B for stake C to all beauties for whom the coin landed heads and received C from OR-beauties for whom the coin landed tails (as he had anticipated). However, he received $2C$ from AND-beauties for whom the coin landed tails. He had erred when using AND-Beauties' choices of wagers to assess their credence in heads because the outcome of the coin toss not only determined whether they won their bet but also the expected cost of the bet. In deciding whether to accept a wager on heads, AND-Beauties must take account not only of the probability of heads but also of the increased cost if the coin lands tails.

OR-beauties and AND-beauties had been offered different wagers on a toss of the same coin. OR-beauties won or lost bets on heads once. By contrast, AND-beauties won bets on heads *once* if the coin landed heads but lost bets on heads *twice* if the coin landed tails. They therefore had more to lose by betting on heads. For each accepted bet, AND-beauties earned $(B - C)$ when the coin fell heads because the bet was placed once but lost $2C$ when the coin fell tails because the bet was placed twice. Bets on heads were better than even money when $(B - C) > 2C$ which is equivalent to $C/B < 1/3$.

If instead, AND-beauties had bet on tails they would have lost C when the coin landed heads but earned $2(B - C)$ when the coin landed tails. Therefore, bets on tails would be better than even money whenever $2(B - C) > C$ which is equivalent to $C/B < 2/3$. Thus, AND-beauties employ different decision rules for bets on heads and tails. Thirders interpret this difference as evidence for unequal credences of heads and tails and would interpret the right-hand sides of $C/B < 1/3$ and $C/B < 2/3$ as credences and the left-hand sides as ratios of stakes to payouts. Halfers deny this interpretation. When an AND-beauty bets on heads, the expected cost is $C^* = 1.5C$ for payout B. Her decision rule $C/B < 1/3$ can be written as $C^*/B < ½$. When an AND-beauty bet on tails, the expected cost is C^* but the payout is $2B$. Her decision rule $C/B < 2/3$ can be written as $C^*/2B < ½$. In the rearranged forms, the left-hand sides represent ratios of stakes to payout and the right-hand sides AND-beauties' consistent credence in heads.

Elga's Sleeping Beauty was an AND-beauty. My calculations of which bets she should accept are not new. They can be found in many analyses that use Dutch Books and the like to probe her 'true' credence in heads. What differs is the interpretation. Hitchcock (2004) concluded that Beauty's credence on waking changed from one-half to one-third because she learned that she was not asleep. (Does a sleeping Sleeping Beauty

have a centered world?) My interpretation is that Beauty's credences do not change. She bets according to her beliefs but these beliefs include an understanding that the stakes depend on the unknown outcome of the coin toss. This diagnosis is not new. Arntzenius (2002) concluded that Beauty's degree of belief in heads should be one-half but that she "should bet at odds that differ from her degrees of belief." Bradley and Leitgeb (2006) similarly distinguished between her credences and fair betting odds. Their analysis parallels my own: what is at stake depends on the toss of the coin.

OR-beauties and AND-beauties differ not because of different beliefs about the frequency of heads but because of different beliefs about how often they must pay if the coin lands tails. Neither OR-beauties nor AND-beauties obtain relevant new information when woken. Both believe that P(heads) = ½. Halfers are vindicated. AND-beauties are fated to choose the same on Monday and Tuesday if the coin lands tails. Therefore, twice as much is at stake if the coin lands tails. As a consequence, H, T_1 and T_2 are given equal weight when deciding whether to bet on heads. From the perspective of an observer of their behavior who cannot ask them to explain their beliefs, AND-beauties behave 'as if' they believed P(heads) = 1/3. Their behavior can be predicted by this belief. In this limited sense, thirders are vindicated.

Hamilton's wager

The preceding analysis was stimulated by thinking about a problem from my doctoral thesis that concerned degrees of relatedness of triploid endosperm (a tissue within seeds). Philosophers may be interested in the parallels. Unfortunately some biological background is necessary. I beg my readers' forbearance.

Inclusive fitness theory was developed to understand fitness trade-offs among kin (Hamilton 1963, 1964). One of the simplest expressions of this theory is known as Hamilton's Rule. This rule-of-thumb predicts the action of natural selection when a gene's expression confers a benefit (B) on one individual's fitness at a cost (C) to another individual's fitness. Natural selection favors the genetic action if $r_b B - r_c C > 0$, where r_b and r_c are measures of the probabilities that the two individuals carry a copy of the gene. Hamilton's Rule can be rewritten as

$$C/B < r_b/r_c \qquad (1)$$

The left-hand side of this inequality is a ratio of the fitness consequences of the action for the two individuals affected and the right-hand side a ratio of their probabilities of carrying recent replicates of the responsible gene. Strictly speaking, Hamilton's Rule is not about particular individuals but about average outcomes of interactions between specified categories of kin. Thus, the relatednesses can be considered to represent the relative frequencies with which a genetic lineage has experienced the costs and benefits of its own action.

Readers will immediately recognize (1) as a restatement of the gambler's decision rule. A gene can be considered to be paying a cost C for a chance of receiving a benefit B. What is uncertain is not whether one class of individuals pays C and another class receives B, but whether the gene placing the bet is present in the relevant individuals, with r_c the frequency of the gene in the class of individuals who have paid the cost and r_b the frequency of the gene in the class of individuals who have received the benefit.

Credences of rational actors are revealed by their acceptance and rejection of wagers. Genes lack beliefs. But the ancestors of present-day genes have passed repeatedly through a sieve that retained variants that made 'better' choices and

discarded those that made 'worse' choices. By this process, present-day genes are expected to behave 'as if' they were rational agents who judge the probability of current events by the past frequency of similar events. Thus, the 'credences' of genetic agents can be inferred from implicit weightings of alternative outcomes in evolutionary games.

Embryos and endosperms

The adjectives haploid, diploid, and triploid refer to nuclei containing one, two, or three copies of each kind of gene. Plants exhibit an alternation of haploid and diploid generations. Offspring may be haploid or diploid and have both haploid and diploid parents. Dad and mom will be used as technical terms to refer to haploid parents and father and mother to refer to diploid parents (Haig 2013). The haploid products of plant meiosis are called spores. Mothers produce megaspores that divide to produce moms. Fathers produce microspores that divide to produce dads. All nuclei of a dad or mom are genetically identical because they are derived from a single product of meiosis. Moms produce eggs and polar nuclei. Dads produce sperm. In most flowering plants, each dad produces two sperm and each mom produces an egg and two polar nuclei. After a process of double fertilization, one of the sperm nuclei of a dad fertilizes the egg nucleus of a mom to form a zygote that develops into a diploid embryo and the other sperm nucleus fuses with both polar nuclei of the mom to form a primary endosperm nucleus that develops into a triploid endosperm (Figure 1). A division of labor during seed development results in the endosperm sacrificing itself for the sake of its twin embryo.

A dad and mom together produce an embryo and endosperm that have identical maternal and identical paternal genomes. The embryo possesses one copy of the

paternal genome for each copy of the maternal genome but the endosperm possesses two copies of the maternal genome for each copy of the paternal genome. The situation of a gene token in endosperm that 'does not know' whether it is maternal (and present in two doses) or paternal (and present in one dose) is analogous to the situation of an AND-beauty who is uncertain whether she bets twice because a coin landed tails or once because the coin landed heads.

What credence should a gene in endosperm have about its parental origin? One-third of randomly chosen tokens from present-day endosperms were inherited from dad and two-thirds from mom. This synchronic observation suggests that endosperm genes should behave 'as if' they had a one-third chance of coming from dad and a two-thirds chance of coming from dad. This is a frequentist view of relatedness. A diachronic perspective suggests a different answer. Any given gene token in endosperm came from mom or dad with equal likelihood. This is a Bayesian view of relatedness (Figure 1).

Each token descends from a 'parent' token from which it received one strand of its double helix. As a token's lineage is traced back into the past, it passes through the bodies of moms and dads in roughly equal proportions (Haig 2012). Therefore, the lineage will have been subject to natural selection half the time as a paternal token and half the time as a maternal token. Tokens of successful lineages might therefore be expected to behave 'as if' maternal and paternal origin were equally likely. The present likelihoods, looking forward, are derived from past frequencies, looking back (Haig 2014).

The question how natural selection 'interprets' the double dose of maternally-derived genes in endosperm relative to the single dose of paternally-derived genes raises similar issues to those debated by halfers and thirders in the Sleeping Beauty

Problem. When an embryo inherits one dose of paternal genes from dad, its associated endosperm also inherits one dose of the same genes. Should not the endosperm's relatedness to dad be the same as the embryo's relatedness to dad? On the other hand, an endosperm inherits a double dose of maternal genes from mom compared to an embryo which inherits a single dose. Should not the greater dilution of paternal genes in endosperms (one-in-three) relative to embryos (one-in-two) mean that the endosperm is less related than the embryo to dad?

Three-card Monte

Questions about the relatedness of an endosperm to its own embryo and to other embryos of the same mother arose in early attempts to apply inclusive fitness theory to seed development. Westoby and Rice (1982) proposed that "alleles in an endosperm are on average three times as likely to reach the next generation through the embryo with which they are associated as through some other embryo … Endosperms would not therefore be selected to acquire extra provisions at the expense of other embryos as strongly as the embryos themselves would be." Queller (1983) similarly concluded that an endosperm would be less assertive in promoting the growth of the embryo in its seed relative to embryos in other seeds than would be the embryo itself. These authors believed an embryo to be less related than its associated endosperm to embryos in other seeds in the ratio one-half to two-thirds.

The conclusions of these authors were soon challenged by Law and Cannings (1984) who found that diploidy versus triploidy made no difference to the assertiveness of endosperm in their population genetic models. This dispute can be considered an argument between endosperm-thirders and endosperm-halfers. A resolution of the

disagreement was proposed by Queller (1984, 1989). Sometimes thirders, sometimes halfers, got the right answer. Who was right depended on details of gene expression. Queller's analysis of the endosperm problem informed my interpretation of the Sleeping Beauty problem.

Consider the relatedness r of an endosperm to its diploid mother and assume that all her embryos are half-sibs. (Readers who consult the primary literature should be aware that the papers cited above consider the relatedness of an endosperm to half-sib embryos rather than diploid mothers. This involves an extra flip of a Mendelian coin giving $r' = 1/3$ as the thirder position and $r' = ¼$ as the halfer position.) The relatedness of an endosperm to its mother corresponds to the probability that a gene in endosperm is of maternal origin. This is analogous to betting on tails in the Sleeping Beauty problem. Betting on heads is analogous to the probability of paternal origin or 'unrelatedness' $(1 - r)$. The Sleeping Beauty problem will be flipped from betting on heads to betting on tails to simplify comparisons. In this conversion, the thirder contention becomes $r = 2/3$, analogous to P(tails) = 2/3, and the halfer contention remains $r = ½$, analogous to P(tails) = ½.

Consider a gene expressed in endosperm that causes a cost C to its associated embryo for a benefit B to its mother when the gene is paternally-derived (single dose), but causes a cost kC to its associated embryo for a benefit kB to its mother when the gene is maternally-derived (double dose). Tokens of paternally-derived genes experience the cost C to its own embryo but do not share in the benefit to the mother whereas tokens of maternally-derived genes experience both the cost kC and the benefit kB. Therefore, a gene will profit on average when $kB - (1 + k)C > 0$ or

$$\frac{C}{B} < \frac{k}{(1+k)} \tag{2a}$$

If the effects of the gene are dominant, (2a) simplifies to $C/B < ½$ because the same costs and benefits are experienced whether the gene is of maternal or paternal origin ($k = 1$). If, on the other hand, the gene has additive effects (proportional to dosage), (2a) simplifies to $C/B < 2/3$ because the costs and benefits when the gene is maternally-derived will be $2C$ and $2B$ ($k = 2$). Thus, a gene expressed in endosperm with dominant effects resembles an OR-beauty in the Sleeping Beauty problem whereas a gene with additive effects resembles an AND-beauty. Queller (1984, 1989) interpreted the right-hand side of (2a) as the relatedness of endosperm to mother. Therefore, he concluded that the "expression-dependent relatedness" was two-thirds for genes with additive effects and one-half for genes with dominant effects.

If a gene expressed in endosperm has dominant effects, then the extra maternal dose has no consequences because a single paternal dose has the same effects as a double maternal dose. By contrast, if the gene has additive effects, then the double maternal dose has twice the influence of the single paternal dose. For this reason, a gene engaged in Hamilton's wager is expected to behave differently depending on whether it has dominant or additive effects.

One of the attractive features of inequality (1) is that it separates a ratio of phenotypic effects (C/B) on the left-hand side from a ratio of genotypic probabilities (r_b/r_c) on the right-hand side. Inequality (2a) loses this pleasing property because a variable that scales costs and benefits (k) appears on the 'relatedness' rather than 'costs and benefits' side of the ledger. The separation of phenotypic effects from genotypic probabilities can be restored by algebraic reshuffling

$$\frac{(1+k)C/2}{kB} < \frac{1}{2} \qquad (2b)$$

The left-hand side of (2b) is now a ratio of stakes (numerator) to payouts (denominator) and the right-hand side is a relatedness of endosperm to mother that is not 'expression-dependent'. Inequalities (2a) and (2b) are algebraically equivalent but (2b) provides greater conceptual clarity. If the right-hand sides of (2a) and (2b) are both interpreted as measures of relatedness then relatedness must mean different things in (2a) and (2b).

If one wishes to interpret evolutionary models in terms of Hamilton's Rule, then inequality (1) must take more complex forms as models become more complex. A theoretician faces an algebraic choice of keeping the left-hand side simple and putting the extra complexity into 'relatedness' or keeping the right-hand side simple and putting the extra complexity into 'costs and benefits'. Queller's (1989) inclusion of factors weighting phenotypic effects into an 'expression-dependent relatedness' is analogous to thirders' inclusion of factors weighting stakes and payoffs into the credence of heads in the Sleeping Beauty problem. Queller preserved the simplicity of 'costs and benefits' at the expense of making 'relatedness' depend on details of gene action.

The status of Hamilton's Rule has recently become a subject of intense dispute within evolutionary biology with passionate critics and defenders (Allen et al. 2013; Liao et al. 2015). Models of the evolution of social interactions are inherently complex. The competing models, if well-formed, should yield similar predictions regardless of their conceptual framework, albeit in different algebraic form. I suspect that much of the heat of this debate arises from alternative algebraic parsings of equations into multivariable 'chunks' that are identified with intuitive concepts such as 'relatedness', 'cost', and 'benefit'. How to parse an equation is often a question of aesthetic preference with

alternative arrangements implicitly defining intuitive concepts in subtlely different ways.

Identity by descent and identity by ascent

In the original Sleeping Beauty problem, there was one Beauty and one coin toss that either fell heads or tails. The toss of a coin was a randomizing device that rendered outcomes uncertain. In the endosperm problem, every endosperm simultaneously contains maternal and paternal gene tokens but each solipsistic token has its own centered world if tokens do not interact. Uncertainty about a token's maternal or paternal origin arises not from randomization but from ignorance. Inclusive fitness theory traditionally assumed that the unpredictable flip of a meiotic coin and ignorance of parental origin were equivalent sources of doubt in determining relatedness.

Consider the relatedness of embryos to their mothers. If one repeatedly sampled gene tokens from current-day mothers and asked whether identical-by-descent (IBD) tokens were present in particular embryos, then the proportion of trials in which the answer was Yes would converge on one-half. This can be considered the synchronic view of relatedness. A diachronic view is useful for understanding the action of past natural selection. As a gene token's lineage is followed back into the past, it passes through bodies of mothers and fathers in roughly equal proportions and is repeatedly present in the germline of mothers interacting with embryos. For each particular embryo, whether IBD tokens were inherited from the mother by the embryo was determined by a flip of a meiotic coin. As the number of coin flips increases, the frequency with which embryos inherited IBD tokens from their mothers should converge on one-half. Thus, from both the synchronic and diachronic views, randomly

selected gene tokens from mothers have probability one-half of IBD tokens in embryos. For these reasons, the relatedness of embryos to mothers is considered one-half.

Now consider the relatedness of mothers to embryos. From the synchronic perspective, half the gene tokens of current-day embryos have identical-by-ascent (IBA) tokens in their mothers. From the diachronic perspective, a token's ancestral lineage will have repeatedly been present in the germline of embryos interacting with their mothers. On average, the embryonic token will have been inherited from the mother in half these interactions. Therefore, as the number of such interactions increases, the frequency with which embryonic tokens interact with IBA tokens in mothers converges on one-half. From both perspectives, randomly-selected gene tokens from embryos have probability one-half of IBA tokens in their mothers. For these reasons, the relatedness of mothers to their embryos has been considered one-half.

Despite the apparent symmetry of the relatedness of embryos to mothers and mothers to embryos, IBD and IBA coefficients of relatedness reflect different sources of uncertainty. The probability of one-half that a maternal gene has IBD tokens in an embryo reflects uncertainty about a flip of a Mendelian coin. A mother possesses two alleles at each locus, one inherited from her mother and one inherited from her father, but only one is transmitted to any particular embryo via the randomizing process of meiosis. By contrast, the probability of one-half that an embryonic gene has IBA tokens in its mother reflects ignorance of parental origin. A randomly chosen gene in an embryo is equally likely to have been inherited from the embryo's mother or father because the embryo receives one gene copy from each.

A general assumption has been that parental origin makes no difference to a gene's effects. Each gene's lineage passes repeatedly through male and female bodies

while its DNA sequence remains unchanged. Therefore, natural selection should act on genes according to their effects averaged across maternal and paternal transmission and genes should behave 'as if' their parental origin is uncertain. However, if genes were to acquire an erasable 'imprint' on passing through male bodies that was reset after passing through female bodies, then the one-half probability that an embryonic gene had IBA tokens in its mother would collapse into a probability of one for genes of maternal origin and zero for genes of paternal origin (Haig 1997, 2000). This would be equivalent to letting Beauty know the outcome of the coin toss before placing her bets.

Some genes, including genes expressed in endosperm, possess locative memories of their parental origin (Haig and Westoby 1989). By processes of natural selection, these imprinted genes should conditionalize their phenotypic effects on parental origin (Haig 2012). These findings have fundamental consequences for how coefficients of relatedness should be calculated in inclusive fitness theory. Organisms no longer possess unified genomes maximizing a unitary fitness but contain maternal and paternal factions with competing agendas (Haig 1997, 2000, 2006).

Conclusions

Beauty's paired wakings on tails and the double dose of genes from mom in endosperm are conceptually similar. The second waking or second dose is a mere doubling of a single draw from the distribution of a random variable. They are duplicate probability. It should be evident that doubling, tripling, or quadrupling the outcome of a single draw does not affect the expected value of the next independent draw of the random variable nor does it change the value of the single draw that has already been made. The probability that a gene came from a mother or father in the previous generation is one-

half as is the probability that a toss of a fair coin is heads no matter how many times the outcomes are repeated.

My analysis has hinged on the conjunctions *and* and *or*. This pivot appeared in the difference between living in a world in which one expects A *and* B while living at temporal locations where one experiences A *or* B. Distinct events that share the same worldly probability are mutually exclusive in locative space. The pivot reappeared in thinking about wagers on duplicate experiences. Beauty's choices changed when she paid for bets in A *and* B rather than A *or* B. The unrecognized conjunction of these two problems of conjunction, and the failure to disentangle them, may partly explain why the Sleeping Beauty problem has been so intractible.

References

Allen B, Nowak MA, Wilson EO (2013) Limitations of inclusive fitness. Proc Natl Acad Sci USA 110:20135–20139.

Bradley D, Leitgeb H. (2006) When betting odds and credences come apart: more worries for Dutch book arguments. Analysis 66:119–127.

Arntzenius F (2002) Reflections on Sleeping Beauty. Analysis 62:53–62.

Bostrom N (2007) Sleeping beauty and self-location: a hybrid model. Synthese 157:59–78.

Briggs R (2010) Putting a value on beauty. Oxf Stud Epistem 3:3–34.

Elga A (2000) Self-locating belief and the Sleeping Beauty problem. Analysis 60:143–147.

Haig D (1997) Parental antagonism, relatedness asymmetries, and genomic imprinting. Proc R Soc B 264:1657–1662.

Haig D (2000) Genomic imprinting, sex-biased dispersal, and social behavior. Ann NY Acad Sci 907:149-163.

Haig D (2006) Intragenomic politics. Cytogenet Genome Res 113:68–74.

Haig D (2012) The strategic gene. Biol Philos 27:461–479

Haig D (2013) Filial mistletoes: the functional morphology of moss sporophytes. Ann Bot 111:337–345.

Haig D (2014) Fighting the good cause: meaning, purpose, difference, and choice. Biol Philos 29:675–697.

Haig D, Westoby M (1989) Parent-specific gene expression and the triploid endosperm. Am Nat 134:147–155.

Hitchcock C (2004) Beauty and the bets. Synthese 139:405–420.

Law R, Cannings C (1984) Genetic analysis of conflicts arising during development of seeds in the Angiospermophyta. Proc R Soc B 221:53–70.

Lewis D (1979) Attitudes *de dicto* and *de re*. Phil Rev 88:513–543.

Lewis D (2001) Sleeping Beauty: reply to Elga. Analysis 61:171–176.

Liao X, Rong S, Queller DC (2015) Relatedness, conflict, and the evolution of eusociality. PLOS Biol 13:e1002098.

Meacham CJG (2008) Sleeping beauty and the dynamics of de se beliefs. Philos Studies 138:245–269.

Queller DC (1983) Kin selection and conflict in seed maturation. J Theor Biol 100:153–172.

Queller DC (1984) Models of kin selection on seed provisioning. Heredity 53:151–165.

Queller DC (1989) Inclusive fitness in a nutshell. Oxford Surv Evol Biol 6:73–109.

Westoby M, Rice B (1982) Evolution of the seed plants and inclusive fitness of plant tissues. Evolution 36:713–724.

Figure 1: A haploid female gametophyte (mom) contributes an egg nucleus to the embryo within a seed and two polar nuclei to the endosperm. These contributions are represented by filled circles. A haploid male gametophyte (dad) contributes a sperm nucleus to both the embryo and the endosperm. These contributions are represented by unfilled circles. What is the 'probability' that a gene in the endosperm comes from dad? From the perspective of an external observer, one-third of the gene tokens in endosperm come from dad, suggesting an answer of one-third. From the perspective of a gene token in endosperm looking backward, it either came from dad or from mom, suggesting an answer of one-half.